REVUE FRANÇAISE D'AUTOMATIQUE, INFORMATIQUE, RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

E. DIDAY

Optimisation en classification automatique et reconnaissance des formes

Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle, tome 6, n° V3 (1972), p. 61-95

<http://www.numdam.org/item?id=RO_1972__6_3_61_0>

R.A.I.R.O.
(6e année, novembre 1972, V-3, p. 61 à 96)

OPTIMISATION EN CLASSIFICATION AUTOMATIQUE ET RECONNAISSANCE DES FORMES

par E. DIDAY (1)

Résumé. — *Les algorithmes actuellement opérationnels, consistant à fournir une « bonne partition » d'un ensemble fini, produisent des solutions dont rien ne permet d'affirmer qu'elles soient optimales. Le principal but de ce texte est une étude synthétique de propriétés d'optimalité dans des espaces formés de partitions d'un ensemble fini. On formalise et on prend pour modèle de cette étude, une famille de techniques particulièrement efficaces du type « nuées dynamiques ». Après avoir développé l'aspect programmation, on illustre les différents résultats par un exemple artificiel et surtout par deux applications concrètes; l'une en géologie minière pour la recherche de familles géographiques de sondages verticaux, l'autre en médecine pour le dépistage de profils biologiques permettant une aide au diagnostic.*

INTRODUCTION

1.1. Le problème

Dans différents domaines scientifiques (médecine, biologie, archéologie, économie etc.), il apparaît fréquemment de vastes ensembles d'objets représentés par un nombre fini de paramètres. Pour le spécialiste, l'obtention des groupements « naturels et homogènes », ainsi que des éléments les « plus représentatifs » d'un tel ensemble constitue une étape importante dans la compréhension de ses données.

Une bonne approche de la solution de ce problème est fournie par les techniques de classification automatique qui consistent à trouver une partition d'un ensemble fini E telle que chaque objet ressemble plus aux objets intérieurs à son groupe, qu'aux objets extérieurs. En termes mathématiques, le problème peut s'énoncer sous l'une des deux formes suivantes; étant donné un certain critère W :

A) Trouver la partition de E qui optimise W.

B) Trouver la partition de E qui optimise W parmi toutes les partitions en K classes.

(1) I.R.I.A. Rocquencourt (France).

La famille de méthodes dont il sera question concerne surtout le problème *B*, mais elle pourra également aider le praticien dans la résolution du problème *C* suivant :

C) *Chercher parmi toutes les partitions en K classes celles dont chaque classe aura le noyau le plus représentatif.* (Un noyau est un groupe de points de la population à classifier) (1).

Dans le paragraphe 1.2. nous donnerons sommairement les principales propriétés des méthodes des nuées dynamiques (2). Cette famille de méthodes servira de modèle au but véritable de cette étude qui sera développé dans le paragraphe 1.3.

1.2. Les méthodes des nuées dynamiques

On se donne une fonction g permettant de transformer une partition de E en un ensemble fini de noyaux et une fonction f permettant le passage de plusieurs noyaux à une partition. Le principe de ces méthodes est simple, il consiste à appliquer de manière alternative les fonctions f et g à partir d'un choix initial de noyaux. Moyennant certaines hypothèses qui seront données, la décroissance du critère W est assurée jusqu'à la convergence. Le formalisme que nous donnons permet d'obtenir de nombreuses variantes de cette technique et notamment, comme cas particuliers, la méthode de Hall et Ball (1965) de Freeman (1969) et de Diday (1970). Nous avons pris cette famille de méthodes comme modèle de notre étude pour de multiples raisons.

a) Elles permettent d'éviter la mise en mémoire du tableau $\frac{N \cdot (N-1)}{2}$ (où $N = \text{card}\,(E)$) des similarités des objets deux à deux. Cela permet le traitement de populations beaucoup plus importantes que par d'autres techniques plus classiques [Sokal et Sneath (1963), Johnson (1967), Roux (1968), Lerman (1970)].

b) Ces techniques sont très rapides, par exemple la variante étudiée dans Diday (1970), permet le traitement sur IBM 360/91 d'une population de 900 objets caractérisés chacun par 35 paramètres en $3\frac{1}{2}$ minutes.

c) Ces techniques ne souffrent pas de l'effet de chaîne [voir Johnson (1967), Zahn (1971)]. Autrement dit, elles n'ont pas tendance à rapprocher deux points éloignés si ces deux points sont liés par une file serrée de points.

d) Il n'est pas nécessaire de définir des seuils arbitraires pour la détermination des classes ni pour l'arrêt du processus [voir Sébestien (1966), Bonner (1964), Hill (1967), etc.].

---

(1) Le problème C est formalisé et un exemple simple est donné en 2.1.
(2) Voir également E. Diday (1970), E. Diday (1971).

L'utilisation des noyaux introduit de plus les avantages suivants :

a) La possibilité de prendre des noyaux de plusieurs éléments de la population permet une reconnaissance plus efficace des formes. On pourra voir, par exemple, dans le chapitre VI de Diday (1971), comment un bon choix de f et g revient à construire des noyaux épousant la structure des formes de E, ce qui ne serait évidemment pas possible si ces noyaux étaient réduits à un seul point. L'utilisation de noyaux permet également un vaste choix des fonctions f et g, par exemple l'utilisation de la distance de Mahalanobis [voir Romeder (1969)] qui n'aurait pas de sens si chaque noyau était réduit à un point unique.

b) En prenant pour noyaux des éléments de la population elle-même, plutôt que des centres de gravités, on évite l'effet artificiel que ces centres peuvent créer (cf. fig. 11). D'un autre côté, pour certains types de données, la notion de centre de gravité peut ne pas avoir de sens.

c) L'utilisation de noyaux permet la réalisation de partitions autour des agglomérations à forte densité en atténuant l'effet des points marginaux (cf. fig. 14 et fig. 15).

d) Signalons enfin, que l'utilisation des noyaux permet de donner des « optimum locaux » au problème C, permettant ainsi d'aider l'utilisateur intéressé par de bons échantillonnages.

1.3. Étude synthétique des solutions obtenues

Toutes les techniques réalisables dont le but est de minimiser le critère W, fournissent des solutions dont rien ne prouve qu'elles soient optimales. Or, les différentes études faites récemment sur l'état actuel des recherches en « clustering » [voir Bolshev (1969), Fisher et Van Ness (1971), Ball (1970), Wattanabé (1971), Cormack (1971)] font ressortir l'inexistence d'étude synthétique des solutions obtenues pour un algorithme donné. C'est à cette étude que ce texte est consacré. Nous nous sommes restreint à un type particulier d'algorithme mais, évidemment, cette analyse pourrait s'étendre à d'autres techniques.

On appellera V_k l'ensemble des solutions possibles. Chaque solution ([1]) obtenue par un algorithme des nuées dynamiques est optimale vis-à-vis d'une certaine partie de V_k qui est une arborescence particulière. Cela conduit à donner une structuration à l'espace V_k. On montre en particulier que, sous certaines hypothèses, cet espace peut être partitionné en un nombre fini d'arborescences qui ont pour racine une solution stable dite « non biaisée » et pour sommets pendants des éléments d'un certain type, appelés « éléments impasses ». On applique les différents résultats obtenus de la manière suivante :

a) On construit une variable aléatoire permettant de se faire une idée réelle de la structure de V_k. On obtient ainsi un invariant intéressant pour de multiples raisons, notamment pour les données évoluant dans le temps et pour comparer l'efficacité des différentes techniques.

b) On définit différents types de « fuzzy-sets » (1) dans E : les formes « fortes » et « faibles » ainsi que les points « charnières ». Bien mieux que l'optimum global, ce sont à notre avis ces « fuzzy-sets » et les optimums locaux obtenus qui fourniront véritablement à l'utilisateur les différentes facettes de la réalité qu'il désire saisir.

c) Nous donnons un nouveau type de techniques permettant par passage d'une arborescence à l'autre, une approche de l'optimum global.

Les exemples d'application qui seront donnés font notamment ressortir l'intérêt des « formes fortes » qui sont un outil d'une grande utilité pour le praticien, en lui permettant d'extraire de sa population les groupes de points les plus significatifs.

Signalons enfin, que nous avons évité les développements théoriques, en nous restreignant aux résultats intéressants pour la compréhension et l'utilisa- informatique des méthodes.

QUELQUES NOTATIONS ET DEFINITIONS

E : l'ensemble des objets à classifier, il sera supposé fini.

$\mathbf{P}(E)$: l'ensemble des parties de E.

$\mathbf{P}_k$: l'ensemble des partitions de E en un nombre $n \leqslant k$ de parties.

$\mathbf{L}_k \subset \{ L = (A_1, ..., A_k) / A_i \subset A \}$ où, selon les cas A représentera E ou $\mathbf{R}^n$ par exemple.

$V_k = \mathbf{L}_k \times \mathbf{P}_k$.

W une application injective : $V_k \to \mathbf{R}^+$.

Un optimum local sur $C \subset V_k$ sera un élément v^* :

$$W(v^*) = \operatorname*{Min}_{v \in C} W(v).$$

Si $C = V_k$ on a un optimum global.

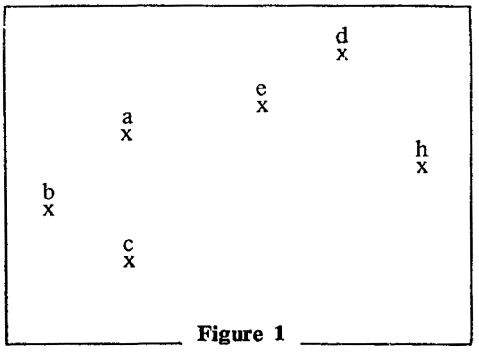


Figure 1

EXEMPLE 1 :

Soit $E = \{ a, b, c, d, e, h \}$ 6 points du plan (voir fig. 1).

W est défini comme suit : soit $v = (L, P)$ où $L = (x_1, x_2) \in \mathbf{L}_2 \equiv E^2$ et $P = (P_1, P_2) \in \mathbf{P}_2$ alors $W(v) = \sum_{i=1}^{2} \sum_{y \in P_i} d(x_i, y)$ où d est la distance Euclidienne.

(1) Voir Zadeh (1965) et Ruspini (1970).

On voit que dans ce cas, l'optimum global est donné par $v^* = (L^*, P^*)$ où $L^* = (b, d)$ et $P^* = \{ \{a, b, c\}, \{e, d, h\} \}$.

EXEMPLE 2 :

Soit E l'ensemble des 17 points indiqués (fig. 2). Prenons

$\mathbf{L}_2 = \{ L = (A_1, A_2)/A_i \subset E, \text{card}(A_1) = 3, \text{card}(A_2) = 2 \}$ et $V_2 = \mathbf{L}_2 \times \mathbf{P}_2$

Choisissons $W(v) = \sum_{i=1}^{2} \sum_{x \in A_i} \sum_{y \in P_i} d(x, y)$ où d est encore la distance Euclidienne.

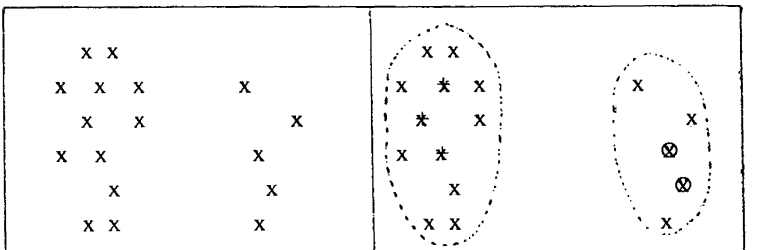


Figure 2 Figure 3

L'optimum global $v^* = (L^*, P^*)$ où $L^* = (A_1^*, A_2^*)$ et $P^* = (P_1^*, P_2^*)$ est donné (fig. 3). Les traits pointillés indiquent les points de E qui constituent P_1^* et P_2^* ; les trois points indiqués par le signe $*$ forment A_1, le signe $\otimes$ sert à représenter les deux points qui constituent A_2.

CONSTRUCTION DE TRIPLETS (f, g, w) PAR LES NUEES DYNAMIQUES

3.1. Formulation générale

Nous noterons $v = (L, P) \in V_k$ où $L \in \mathbf{L}_k$: $L = (A_1, ..., A_k)$ avec $A_i \subset A$ et $P \in \mathbf{P}_k$: $P = (P_1, ..., P_k)$ où les P_i sont les classes de la partition P de E. On se donne également les quatre applications suivantes :

$D : E \times \mathbf{P}(E) \to \mathbf{R}$ qui dans la pratique exprimera la similarité d'un élément de E avec une partie de E.

$R : E \times T \times \mathbf{P}_k \to \mathbf{R}^+$ (où T est l'ensemble des entiers compris entre 1 et k). Cette application servira à agréger et à écarter les classes entre elles. On peut prendre par exemple $R(x, i, P) = D(x, P_i)$. On aurait pu également définir $R : E \times T \times V_k \to \mathbf{R}^+$ (cf. [9]); comme le montre l'exemple ci-dessous, cette

définition de R peut rendre des services, cependant, afin de simplifier nous nous restreindrons dans tout ce texte à la première définition.

EXEMPLE : $R(x, i, v) = \dfrac{D(x, P_i)}{(D(x, A_i))^{\frac{1}{n}}}$

Plus n est grand, moins les éléments des noyaux seront dispersés dans chacune des classes obtenues et plus ils exprimeront le squelette des formes qu'ils déterminent (cf. [10] chap. VI, un exemple de cas limite avec $n \to \infty$). Par un choix adéquat de n l'utilisateur pourra ainsi tenter d'obtenir des noyaux dont la distribution soit une bonne image de la distribution des classes qui leur correspond.

Le triplet (f, g, W) est construit comme suit :

le critère $W : V_k \to \mathbf{R}^+ : v = (L, P) \Rightarrow W(v) = \sum_{i=1}^{k} \sum_{x \in A_i} R(x, i, P)$

l'application $f : \mathbf{L}_k \to \mathbf{P}_k$ est définie par : $f(L) = P$

avec $P_i = \{ x \in E / D(x, A_i) \leqslant D(x, A_j)$ pour $j \neq i \}$, en cas d'égalité on affecte x à la partie de plus petit indice.

l'application $g : P_k \to L_k$ est définie par : $g(P) = L$

avec $A_i =$ les n_i éléments $a \in A$ qui minimisent $R(a, i, P)$. La valeur des n_i dépendra de la variante choisie (cf. 3.2.).

Dans [9] nous avons pris l'habitude d'appeler A_i « noyau » de la $i^{\text{ième}}$ classe, et « étalons » les éléments qui le constituent.

REMARQUE : si $R : E \times T \times V_k \to \mathbf{L}_k$, il faut choisir $g : V_k \to \mathbf{L}_k$.

L'algorithme des nuées dynamiques consiste à appliquer alternativement la fonction f puis la fonction g sur le résultat obtenu et cela à partir de $L^{(0)} \in \mathbf{L}_k$ estimé ou tiré au hasard.

3.2. Les différentes variantes et intérêt comparé

Nous ne prétendons pas exposer ici toutes les variantes possibles; nous exposerons celles qui ont paru les plus intéressantes, en faisant simplement varier le choix de g et de R (laissant au lecteur le loisir d'en imaginer d'autres !).

a) Pour cette variante on a : $A \equiv \mathbf{R}^n$; $n_i = 1 \; \forall \, i$;

si on prend de plus $R(x, i, P) = D(x, P_i)$, $g(P) = L$ est tel que A_i soit le centre de gravité (1) de P_i au sens de D.

Hall et Ball proposent une méthode de ce type dans [13].

(1) x est appelé centre de gravité de P_i au sens de D si $D(x, P_i) = \inf_{x \in \mathbf{R}^n} D(x, P_i)$.

b) $A \equiv E$ et $n_i = \text{card } (F_i)$ où :

$F_i = \{ x \in E / R(x, i, P) \leqslant R(x, j, P) \forall j \neq i \}$, si $i < j$ et $R(x, i, P) = R(x, j, P)$ on affecte x à F_i. On voit alors que les A_i sont identiques aux F_i et constituent une partition de E.

On trouvera une étude approfondie de ce cas dans [9] (qui est une généralisation de la méthode proposée par Freeman dans [12] où $\mathbf{L}_k \equiv \mathbf{P}$ et g est remplacée par f. Notons qu'une variante intéressante de cette méthode consisterait à choisir $Vi \in \{1, 2, ..., k\}$ $n_i = \alpha \text{ card } (F_i)$ avec $\alpha = \frac{1}{3}$ par exemple.

c) $A \equiv E$ et n_i fixé une fois pour toute $\forall i \in \{1, ..., k\}$; n_i sera choisi par l'utilisateur s'il a quelque idée du contenu de ses données, sinon il pourra prendre $n_i = \dfrac{\alpha \text{ card } E}{k}$ pour tout i. (Voir [9] et [10].)

d) $A \equiv E$, n_i fixé ou égal à $\alpha \text{ card } P_i$ avec $0 < \alpha < 1$; on définit A_i comme étant les n_i éléments de P_i qui minimisent $R(x, i, P)$. Quand n_i est fixé et dans le cas où le nombre d'étalons d'un noyau devient supérieur au nombre d'éléments de la classe correspondante, on prendra par exemple, $n_i = \text{card } P_i$ s'il s'agit de la classe P_i.

REMARQUE

Dans le cas où l'utilisateur désire obtenir des classes empiétantes, il lui suffit de prendre $\alpha > 1$ dans les variantes *b*) et *c*).

On peut construire des méthodes mélangeant les variantes :

On pourrait ainsi commencer par une variante du type *c*) pour localiser les formes, puis terminer par une variante du type *d*) pour que les A_i donnent assurément les éléments les plus représentatifs de la classe P_i. Dans tous les cas où l'utilisateur a besoin de définir des contraintes sur les noyaux, on choisira A de manière à ce que les éléments des noyaux satisfassent ces contraintes.

EXEMPLES :

1) faire une typologie d'un ensemble d'entreprises mais en imposant aux noyaux de n'être formé que d'entreprises modèles;

2) faire une typologie sur un ensemble de formes en imposant aux noyaux d'être pris parmi un ensemble de formes types.

Dans le cas où $K = 1$ on peut utiliser la variante suivante qui permet d'obtenir des étalons aux endroits à forte densité :

$L^{(0)} = A_1^{(0)} = n$ points tirés au hasard dans E.

$P^{(q)} =$ les m points de E les plus proches de $A_1^{(q-1)}$ au sens de D avec $m > n$ (par exemple $m = n + 1$).

$A_1^{(q)} =$ les n points de E les plus proches de $P^{(q)}$ au sens de R.

Cette technique peut donner à l'utilisateur une idée a priori du nombre de classes de E.

Comparaison des différentes variantes : par rapport a la variante a), les variantes b), c) et d) ont l'avantage d'atténuer l'effet artificiel créé par des centres de gravité en utilisant des noyaux d'éléments de la population elle-même.

La variante b) a l'avantage par rapport à c) de ne pas nécessiter l'introduction des paramètres n_i, cependant elle a une forte tendance à osciller au lieu de converger, elle donne plus d'importance aux éléments marginaux puisque ses noyaux recouvrent E, alors que les noyaux de la variante c) ne tiennent compte que des éléments les plus représentatifs; de plus, elle nécessite beaucoup plus de calculs et de place mémoire qu'une utilisation de c) avec $\Sigma n_i < <$ card E.

La variante d) permet d'assurer la représentativité des noyaux vis-à-vis de leur classe et réduit les calculs et la place mémoire; cependant le choix des noyaux étant moins vaste à chaque itération (puisqu'on astreint les étalons à n'appartenir qu'à l'une des classes précédentes) elles peut donner des classes moins pertinentes que pour la variante c).

3.3. Construction de triplets rendant la suite u_n décroissante

Définitions des suites u_n et v_n :

Soit h l'application $V_k \to V_k$ telle que :

$(v = L,P) \in V_k \Rightarrow h(v) = (g(P), f(g(P)))$.

Une suite $\{ v_n \}$ est définie par v_0 et $v_{n+1} = h(v_n)$.

Une suite $\{ u_n \}$ est définie à partir d'une suite $\{ v_n \}$ par $u_n = W(v_n)$.

Définition de S :

Soit $S : \mathbf{L}_k \times \mathbf{L}_k \to \mathbf{R}^+$:

$$S(L, M) = \sum_{i=1}^{k} \sum_{x \in A_i} R(x, i, Q) \qquad \text{où} \qquad Q = f(M).$$

Définition d'une fonction carrée (1)

On dira que R est carrée si :

$$S(L, M) \leqslant S(M, M) \Rightarrow S(L, L) \leqslant S(L, M).$$

Théorème 1 :

Si R est carrée le triplet (f, g, W) rend la suite u_n décroissante pour les variantes où le nombre d'étalons par noyau est fixé.

---

(1) Nous avons exhibé dans [9] un exemple de fonction R carrée.

Démonstration

Étant donnée la suite $v_n = (L^{(n)}, P^{(n)})$ on a :

$$W(v_n) = \sum_{i=1}^{k} \sum_{x \in Ai^{(n)}} R(x, i, P^{(n)}) \qquad \text{où} \qquad L^{(n)} = (A_1^{(n)}, \ldots, A_k^{(n)})$$

d'où : $u_n = W(v_n) = S(L^{(n)}, L^{(n)})$.

Posons $z_n = S(L^{(n+1)}, L^{(n)})$.

Si R est carrée et si $z_n \leqslant u_n$ on a nécessairement :

$$S(L^{(n+1)}, L^{(n)}) \leqslant S(L^{(n)}, L^{(n)}) \rightarrow S(L^{(n+1)}, L^{(n+1)}) \leqslant S(^{(n+1)}, L^{(n)}) \rightarrow u_{n+1} \leqslant z_n.$$

Montrons que $z_n \leqslant u_n$; en effet :

$$u_n = \sum_{i=1}^{k} \sum_{x \in A_i^{(n)}} R(x, i, P^{(n)}) \geqslant \sum_{i=1}^{k} \sum_{x \in A_i^{(n+1)}} R(x, i, P^{(n)}) = z_n$$

par construction même de $A_i^{(n+1)}$.

On voit ici l'intérêt de fixer le nombre d'étalons par noyau, car cette dernière inégalité n'est pas nécessairement vérifiée dans le cas de la variante *b*).

On a finalement montré que : $u_{n+1} \leqslant z_n$ et $z_n \leqslant u_n$ d'où $u_{n+1} \leqslant u_n$.

c.q.f.d.

N. B. Dans toute la suite on se restreindra au cas où le nombre d'étalons par noyau est fixé.

STRUCTURATION DE L_k, P_k, V_k ET PROPRIETES D'OPTIMALITE

Considérons le graphe $\Gamma = (V_k, h)$. Il apparaît alors, des éléments particuliers dans V_k :

1) *Les éléments non biaisés*

Les propriétés suivantes sont équivalentes et caractérisent un élément non biaisé (1) $v = (L, P) \in V_k$.

a) v est racine d'une arborescence bouclée de Γ.

b) v est un point fixe de h.

c) $L = g(P)$, $f(L) = P$.

(1) Cette appelation vient du fait que les noyaux correspondants à un tel élément sont au centre (au sens de g) de la classe qu'ils déterminent au sens de f.

Les propriétés *d*) (resp. *e*), permettent de caractériser les éléments non biaisés de $\mathbf{L}_k$ (resp. $\mathbf{P}_k$).

d) $g(f(L)) = L$.

e) $f(g(P)) = P$.

2) *Les éléments impasses*

Les propriétés a) et b) suivantes sont équivalentes et caractérisent un élément impasse $v = (L, P) \in V_k$.

a) v est un sommet pendant de Γ.

b) $P \neq f(L)$ ou $f^{-1}(g^{-1}(L)) = \varnothing$

Signalons que les propriétés *c* (resp. *d*) suivantes, permettent de caractériser les éléments impasses de $\mathbf{L}_k$ (resp. $\mathbf{P}_k$).

c) $g^{-1}(L) = \varnothing$ ou $f^{-1}(g^{-1}(L)) = \varnothing$.

d) $f^{-1}(P) = \varnothing$ ou $g^{-1}(f^{-1}(P)) = \varnothing$ ou $f^{-1}(g^{-1}(f^{-1}(P))) = \varnothing$.

Le théorème suivant se déduit immédiatement des définitions de la proposition 2 (cf. annexe 1) et du théorème 1.

Théorème 2 :

Si R est carrée alors :

a) Chaque composante connexe de $\Gamma = (V_k, h)$ est une arborescence bouclée.

b) Il existe dans V_k au moins un élément non-biaisé.

c) Si un élément non-biaisé $v \in V_k$ est sommet d'une arborescence C, alors v est un optimum local (1) vis-à-vis de l'ensemble des sommets de C.

d) Si $w \in V_k$ n'est pas un élément non-biaisé alors w appartient à une arborescence bouclée de racine w^*, et $W(w) > W(w^*)$. Et l'optimum global est un élément non-biaisé.

REMARQUE :

On a deux énoncés équivalents de ce théorème, en remplaçant partout V_k par $\mathbf{L}_k$, puis par $\mathbf{P}_k$. Il suffit pour cela d'utiliser les fonctions $\varphi_1 : V_k \to \mathbf{L}_k$ et $\varphi_2 : V_k \to \mathbf{P}_k$ telles que :

$$\text{si } v = (L, P) \text{ alors } \varphi_1(v) = L \text{ et } \varphi_2(v) = P.$$

On voit d'après ce théorème que dans le cas où R est carrée, il existe trois types d'éléments dans V_k (et de même dans $\mathbf{L}_k$ et $\mathbf{P}_k$) : les éléments impasses, les éléments non-biaisés et les éléments restants qui seront dit « biaisés ».

(1) On peut montrer que v, est un optimum local pour une certaine topologie basée sur la différence symétrique. On peut montrer également que l'algorithme de Mac Queen [28] ne converge pas nécessairement vers une solution non-biaisée.

Si R n'est pas carrée, on est dans le cas de la proposition 1 ([1]), il y a des circuits dans V_k (et de même dans $\mathbf{L}_k$ et $\mathbf{P}_k$); autrement dit on peut trouver des suites $\{ v_n \}$ pour lesquelles il existe $N > 1$ tel que $v_0 = v_N$. Les éléments de ces circuits constituent donc un quatrième type d'éléments de V_k, $\mathbf{L}_k$ ou $\mathbf{P}_k$.

Exemple de différents types d'éléments de $\mathbf{L}_k$:

Reportons nous au cas de la figure 1 et prenons le triplet (f, g, W) des nuées dynamiques (cf. 3.1.), avec $k = 2, n_1 = n_2 = 1, A \equiv E \equiv \{a, b, c, d, e, h\}$,

$$\mathbf{L}_2 \equiv A \times A, D(x, Y) = \sum_{y \in Y} \mathrm{d}(x, y)$$

où d est la distance euclidienne et $R(x, i, P) = D(x, P_i)$.

Elément impasses :

Il n'existe pas deux points de E permettant d'engendrer à l'aide de f la partition $P = (P_1, P_2)$ où :

$$P_1 = \{ a, d, h \} \text{ et } P_2 = \{ b, c, e \}.$$

Ainsi $f^{-1}(P) = \varnothing$, comme $L = (e, c) = g(P)$ on peut dire que L est un élément impasse. Un autre exemple d'élément impasse est le point $L = (b, h)$ car $g^{-1}(L) = \varnothing$.

Elément non-biaisé :

$L = (b, p) \in \mathbf{P}_2$ est un élément non-biaisé car on voit simplement d'une part que $f(L) = P = (P_1, P_2)$ avec $P_1 = \{a, b, c\}$ et $P_2 = \{d, e, h\}$ et d'autre part que $g(P) = L = (b, d)$.

Élément biaisé :

$L = (c, e)$ est biaisé car $g(f(L)) = (b, d) \neq L$.

Nous avons utilisé l'appelation « élément biaisé » car un tel élément (c, e) n'est pas le plus proche de la partition qu'il engendre, contrairement à (b, d) qui est non-biaisé car il vérifie cette propriété.

RECHERCHE D'INVARIANTS

5.1. Mesure des arborescences

On supposera d'abord que le triplet (f, g, W) rend u_n décroissant $\forall u_0$ [autrement dit que $\forall x \in V_k$, $W(h(x)) < W(x)$]. L'espace probabilisé $(\Omega, \mathcal{A}, P)$ des familles d'arborescences bouclées est définit comme suit :

$$\Omega = V_k.$$

(1) Cf. annexe 1.

$\mathcal{A}$ = l'algèbre engendrée par la partition de Ω en arborescences bouclées (c'est-à-dire ensemble des parties de Ω qui sont réunions d'arborescences bouclées).

$P : \mathcal{A} \to [0, 1]$ est telle que si $C \in \mathcal{A}$ est réunion de n arborescences bouclées $C_1, ..., C_n$ alors

$$P(C) = \frac{1}{\text{card } (\Omega)} \sum_{i=1}^{n} \text{card } C_i.$$

La variable aléatoire X (dite des familles d'arborescences) de $(\Omega, \mathcal{A}, P)$ dans (R, B) où B est la tribu borélienne, est l'application $\Omega \to \mathbf{R}$ telle que $X(v) = W(w)$ où w est l'élément non biaisé de l'arborescence bouclée contenant v. X est bien une variable aléatoire car si $I \in B$, $X^{-1}(I)$ est la réunion d'arborescences de V_k ayant pour sommets les éléments v tels que $X(v) \in I$. La fonction de répartition $F(x) = \text{pr}[X < x]$ exprime la probabilité d'obtenir un élément $v \in V_k$ dans une arborescence bouclée ou une boucle contenant un élément non-biaisé w tel que $W(w) < x$. On donnera en 7.1 un exemple de fonction de répartition empirique correspondant à un n-échantillon de V_k. Dans le cas où $\exists x$: $W(f(x)) > W(x)$ (autrement dit, si on ne suppose plus la suite u_n décroissante $\forall u_0$), on peut également définir une variable aléatoire des composantes connexes de V_k. La variable aléatoire de $(V_k, \mathcal{A}, P)$ dans $(\mathbf{R}, B)$ est telle que $X(v) = \inf_{y \in C} W(y)$ où C est la partie connexe de V_k à laquelle appartient v.

L'introduction de ces variables aléatoires permet de se faire une idée du nombre de composantes connexes et de leur taille respective, grâce aux fonctions de répartitions empiriques. Cela donne également un outil de comparaison des différentes techniques, la meilleure étant celle pour laquelle les racines des arborescences de plus grande taille correspondent aux plus petites valeurs prises par W (voir 7.1).

5.2. Formes fortes, fuzzy-sets et information

5.2.1. Caractérisation des différents types de formes

Soient $C_1, ..., C_n$, n parties connexes du graphe (V_k, h) et $C = C_1 \times C_2 \times ... \times C_n$; on définit l'application $Z : C \to \mathbf{R}^+$ par

$$Z(V) = W(v_1) + ... + W(v_n) \text{ où : } V = (v_1, ..., v_n) \in C \text{ et } v_i \in C_i.$$

Soit $V^* : Z(V^*) = \underset{V \in C}{\text{Min}} Z(V)$. Soit $V^* = (v_1^*, ..., v_n^*)$ et $v_1^* = (L^{i*}, P^{i*})$.

(Si R est carrée, C_i est une arborescence bouclée ou une boucle et v_i^* est l'élément non-biaisé de C_i.) Notons P_j^i la $j^{\text{ième}}$ classe de la partition P^i.

Soit H l'application $E \Rightarrow \mathbf{N}^n$ qui, à chaque élément $x \in E$ fait correspondre le vecteur $(\alpha_1, ..., \alpha_n)$ où α_i est le numéro de la classe ou apparaît l'élément x dans P_i^*. Soit $H(y) = (\beta_1, ..., \beta_n)$ et $\delta(x_,, y)$ le nombre d'indices i pour $i = 1, 2, ..., n$ tels que $\alpha_i - \beta_i = 0$.

Soient F_n et F_1 deux applications multivoques définies sur E telles que

$$F_n(x) = \{y \in E / \delta(x, y) = n\} \quad \text{et} \quad F_1(x) = \{y \in E / \delta(x, y) \geqslant 1\}.$$

Définition des formes fortes (1) :

Les propriétés suivantes sont équivalentes et caractérisent la partition P^* de E dont chaque classe est une forme forte.

1) $P^* = P^{1*} \cap P^{2*} \cap \ldots \cap P^{n*}$.

2) P^* est la moins fine (2) des partitions qui sont plus fines que $P^{1*}, \ldots, P^{n*}$.

3) P^* est la partition définie par l'espace quotient E/H.

4) P^* est la partition définie par les parties connexes du graphe $\Gamma_n = (E, F_n)$.

Définition des formes faibles:

Les propriétés suivantes sont équivalentes (3) et caractérisent la partition Q^* de E dont chaque classe est une forme faible.

1) Q^* est la plus fine des partitions qui sont moins fines que $P^{1*}, \ldots, P^{n*}$.

2) Q^* est la partition définie par l'ensemble des parties connexes de $\Gamma_1 = (E, F_1)$.

Plus généralement, si on pose $F_p(x) = \{y \in E / \delta(x, y) \geqslant p\}$ et $\Gamma_p = (E, F_p)$, l'ensemble des parties connexes de Γ_p pour $p = 0, 1, 2, \ldots, n$ constitue une hiérarchie. Cette hiérarchie induit l'ultramétrique sous dominante de la différence symétrique (cf. annexe 3).

REMARQUE :

On voit, d'après ces définitions que P^* est une partition plus fine que Q^*.

Définitions des points charnières et des points isolés:

On les caractérise par le fait que ce sont les formes fortes réduites à un seul point. Ils se distinguent par la propriété suivante :

un point $a \in E$ est isolé si $\delta(a, x) = o \forall x \in E$.

un point $a \in E$ est charnière si $\exists x \in E : o < \delta(a, x) < n$.

5.2.2. Fuzzy sets (4)

L'intérêt des « fuzzy sets » de Zadeh que nous introduisons ici est qu'ils permettent :

a) d'obtenir de nouvelles formes à la suite d'opérations ensemblistes sur les formes fortes (réunion, intersection, etc.);

---

(1) L'intersection de deux partitions est l'ensemble des parties obtenues en prenant l'intersection de chaque classe de l'une par toutes les classes de l'autre.
(2) Une partition P est dite plus fine qu'une partition P' de E si toute classe de P' est union de classe de P.
(3) Pour la démonstration de cette équivalence cf. annexe 2.
(4) Voir [20] et [26].

b) de caractériser ces nouvelles formes sans avoir besoin de définir des profils types (par des calculs de moyenne par exemple) ni même de connaître les éléments qui les constituent;

c) d'utiliser au maximum l'information apportée par le tableau des formes fortes.

Chaque forme forte A peut être considérée comme un « fuzzy-set » caractérisé par l'application $h_A : E \to [0, 1]$ telle que $h_A(x) = \dfrac{\delta(x, a)}{n}$ où $a \in A$. On voit d'après la définition (3e propriété) que $h_A(a) = 1 \ \forall a \in A$. On peut utiliser h_A pour avoir une idée du degré de ressemblance avec A d'un point charnière ou d'une autre forme forte. On peut également utiliser l'application $F : \mathcal{F} \to [0, 1]$ où $\mathcal{F}$ est l'ensemble des formes faibles de E et

$$F(B) = \frac{1}{\text{card } (B)} \sum_{x \in B} \left(\frac{1}{m} \sum_{j=1}^{m} h_{Aj}(x) \right)$$

où les Aj sont les m formes fortes qui constituent B. Cette application F exprime le degré de faiblesse de B car plus les formes fortes Aj sont dissemblables plus $F(B)$ sera petit. La plus grande valeur de F est 1, c'est-à-dire quand B est une forme forte.

5.2.3. Stabilité des formes fortes et information

Si le nombre de classes demandées est K et si n est le nombre d'optimum locaux obtenus, il est clair que $\forall x \in E$, $H(x)$ peut prendre $(k)^n$ valeurs, cela souligne la cohésion des éléments d'une forme forte A puisque si x et $y \in A$ on a $H(x) = H(y)$. Cependant l'utilisateur, désireux d'avoir une assurance supplémentaire en ce qui concerne la cohésion et la stabilité des formes fortes, peut utiliser l'information apportée par l'augmentation de n. Considérons les classes $P_1^{q*}, \ldots, P_k^{q*}$ de la partition P_q^*, et soient $A_1, \ldots, Am$ les formes fortes obtenues pour $n = q - 1$; soit $P(j/i) = \dfrac{1}{\text{card } A_i} \text{card } (A_i \cap P_j^{q*})$, cette quantité exprime la probabilité pour un élément d'être dans P_j^{q*} sachant qu'il est dans A_i.

On peut maintenant mesurer l'information apportée par P_q^* connaissant $A_1, \ldots, A_m$:

$$I(P^{q*}/P^{1*}, \ldots, P^{q-1*}) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} P(j/i) \log_k P(j/i).$$

Si la partition des formes fortes $A_1, \ldots, A_m$ est plus fine que la partition P^{q*} l'information apportée par P^{q*} est nulle puisque $P(j/i) = 1$ ou 0. Si chacune des formes fortes A_j est répartie de manière égale dans chaque classe P_j^{q*} pour $j = 1, \ldots, k$ alors l'information apportée est maximum et vaut 1.

L'invariance des formes fortes est donc assurée pour $n = q$ si $\forall n > q$ on a $I(P^{n*}/P^{1*}, ..., P^{n-1*}) = 0$. Ainsi, sur les données de Ruspini (cf. 7.1) on s'aperçoit qu'au-delà de $n = 4$ l'information nouvelle reste généralement nulle.

Signalons, enfin, que le nombre $J(k) = \sum_{q=1}^{n} I(P^{q*}/P^{1*}, ..., P^{q-1*})$ peut donner une idée sur la valeur du choix du nombre de classes k demandé; la plus petite valeur de $J(k)$ correspondant au meilleur choix de k.

5.3. Optimum global de V_k

Théorème 3 :

Si les hypothèses suivantes sont vérifiées :

1) R est carrée.

2) $I(P^{n*}/P^{1*}, ..., P^{n-1*}) = 0 \; \forall \; n : q < n < N$ où N est le nombre d'arborescences bouclées.

Alors, la partition des formes fortes $P^{1*} \cap ... \cap P^{q*}$ est plus fine que la partition correspondant à l'optimum global v^* de V_k.

Démonstration :

L'optimum global v^* est racine d'une arborescence bouclée ou d'une boucle puisque d'après 1) on peut utiliser le théorème 2.

Soit P^{i*} la position correspondant à v^* si $i \leqslant q$, P^{i*} est moins fine que la partition des formes fortes P^* car $P^* = P^{1*} \cap ... \cap P^{q*}$. Si $i > q$ l'information apportée par P^{i*} est nulle et donc P^{i*} est moins fine que P^*.

c.q.f.d.

Ainsi, sous les hypothèses de ce théorème, chaque classe de la partition correspondant à l'optimum global est une réunion de formes fortes qui doivent être proches. Pour représenter cette proximité on peut par exemple utiliser une analyse factorielle du triple ou un « minimum spaning tree » (cf. [23]), sur le tableau $T(i, j) = h_{A_i}(a_j)$ où A_i est la $i^{\text{ième}}$ forme forte et a_j un élément de la $j^{\text{ième}}$ forme forte ou encore la méthode des connexités descendantes (cf. annexe 3). Cette méthode permet (sous les hypothèses du théorème 3) une bonne approche de l'optimum global.

Remarquons que la 2e hypothèse est vérifiée pour q d'autant plus petit qu'il existe effectivement k formes dans la population E.

5.4. Approche de l'optimum global par changement d'arborescences

Il s'agit de construire à l'aide de deux éléments non biaisés v_1 et v_2 de V_k un troisième élément non biaisé v_3 qui améliore le critère. Nous supposerons R carrée.

Nous noterons

$$v^i = (L^i, P^i) \in V_k \text{ avec } L^i = (L^i_1, ..., L^i_k) \text{ et } P^i = (P^i_1, ..., P^i_k);$$

$$v^i_j = (L^i_j, P^i_j).$$

Nous supposons que W est additive ce qui est souvent le cas dans la pratique) autrement dit, qu'il existe une application $z : \mathbf{P}(E) \times \mathbf{P}(E) \to \mathbf{R}^+$ telle que :

$$W(v^i) = \sum_{j=1}^{k} z(v^i_j).$$

Supposons que v^1 et v^2 soient deux solutions non biaisées et soit $\{ v^i_{j_1}, ..., v^i_{j_k} \}$, les k plus petites valeurs prises par $z(x)$ avec $x \in \{ v^i_j / i = 1, 2 \text{ et } j = 1, 2, ... k \}$. Notons $P = (P^i_{j_1}, ..., P^i_{j_k})$ et $L = (L^i_{j_1}, ..., L^i_{j_k})$. On montre alors facilement la proposition suivante.

Proposition :

Si $L \neq L^1$, $L \neq L^2$ et $P \in \mathbf{P}_k$ alors l'arborescence bouclée contenant)]. $v = (L, P)$ à pour racine un élément non biaisé v_3 : $W(v_3) < \inf [W(v_1), W(v_2$

PROGRAMMATION DU TABLEAU DES FORMES FORTES ET INTERPRETATION HEURISTIQUE

En ce qui concerne la programmation des méthodes des nuées dynamiques, de larges développements pourront être trouvés dans [8], [9], [10]. Nous signalerons donc simplement l'apport qui a été fait depuis, par la sortie automatique des formes « fortes » et « faibles ». Le programme donne en sortie un tableau [dit des formes fortes (1)] représentant les différents types de formes; la première colonne de ce tableau (cf. tableau 1) donne le nom de chaque élément de E dans un ordre tel qu'à la suite de chaque élément x apparaît l'élément y rendant $\delta(x, y)$ minimum (2). On trouve sur chaque ligne le nom de l'élément x suivi

(1) Par opposition au tableau des formes faibles qui aurait en général une structure différente. Signalons toutefois qu'il est possible de construire un tableau respectant simultanément la structure en formes fortes et en formes faibles.

(2) Nous revenons ici aux notations données en 5.2.1.

TABLEAU 1. — *Formes fortes pour les données de Ruspini*

NUMÉRO DES POINTS	SOLUTIONS OBTENUES						
	1re	2e	3e	4e	5e	δ_{ij}	
1	1	3	3	1	4	0	
2	1	3	3	1	4	0	
3	1	3	3	1	4	0	A_1
5	1	3	3	1	4	0	
6	1	3	3	1	4	0	
9	1	3	3	1	4	0	
10	1	3	3	1	4	0	
8	1	3	3	1	6	1	
4	1	3	3	1	6	0	
7	1	3	3	1	6	0	A_2 B_1
11	1	3	3	1	6	0	
12	1	3	3	1	6	0	
13	1	3	3	1	6	0	
14	1	3	3	1	5	1	
15	1	3	3	1	5	0	
16	1	3	3	1	5	0	
17	1	3	3	1	5	0	A_3
18	1	3	3	1	5	0	
19	1	3	3	1	5	0	
20	1	3	3	1	5	0	
21	2	2	1	2	2	5	
22	2	2	1	2	2	0	
23	2	2	1	2	2	0	
24	2	2	1	2	2	0	
25	2	2	1	2	2	0	
26	2	2	1	2	2	0	
27	2	2	1	2	2	0	
28	2	2	1	2	2	0	
29	2	2	1	2	2	0	
30	2	2	1	2	2	0	
31	2	2	1	2	2	0	
32	2	2	1	2	2	0	A_4
33	2	2	1	2	2	0	
34	2	2	1	2	2	0	
35	2	2	1	2	2	0	
36	2	2	1	2	2	0	
37	2	2	1	2	2	0	
$U =$	1.2135	1.2135	4.8910	1.2135	1.0000		

n° novembre 1972, V-3.

TABLEAU 1 *(suite)*. — *Formes fortes pour les données de Ruspini*

NUMÉRO DES POINTS	SOLUTIONS OBTENUES							
	1re	2e	3e	4e	5e	δ_{ij}		
38	2	2	1	2	2	0		
39	2	2	1	2	2	0		
40	2	2	1	2	2	0		
41	2	2	1	2	2	0		B_2
42	2	2	1	2	2	0		
43	2	2	1	2	2	0		
44	3	4	1	4	1	4		
45	3	4	1	4	1	0		
46	3	4	1	4	1	0		
47	3	4	1	4	1	0		
48	3	4	1	4	1	0		
49	3	4	1	4	1	0		
50	3	4	1	4	1	0		
51	3	4	1	4	1	0		
52	3	4	1	4	1	0	A_5	
53	3	4	1	4	1	0		
54	3	4	1	4	1	0		
55	3	4	1	4	1	0		
56	3	4	1	4	1	0		
57	3	4	1	4	1	0		
58	3	4	1	4	1	0		
59	3	4	1	4	1	0		
60	3	4	1	4	1	0		
61	4	1	2	3	3	5		
62	4	1	2	3	3	0		
63	4	1	2	3	3	0		
64	4	1	2	3	3	0		
65	4	1	2	3	3	0		
66	4	1	2	3	3	0		
67	4	1	2	3	3	0		
68	4	1	2	3	3	0	A_6	B_3
69	4	1	2	3	3	0		
70	4	1	2	3	3	0		
71	4	1	2	3	3	0		
72	4	1	2	3	3	0		
73	4	1	2	3	3	0		
74	4	1	2	3	3	0		
75	4	1	2	3	3	0		
$U =$	1.2135	1.2135	4.8910	1.2135	1.0000			

des valeurs $\alpha_1, ..., \alpha_n$ qui décrivent le vecteur $H(x)$. La valeur $\Delta(x, y) = n\text{-}\delta(x, y)$ correspondant à deux éléments consécutifs x, y est donnée en dernière colonne et permet une détection aisée des formes fortes et faibles : si $\Delta(x, y) = n$ cela signifie que x est le dernier élément d'une forme faible (c'est-à-dire d'une partie connexe de Γ_1); si $\Delta(x, y) = 0$ cela signifie que x et y font partie de la même forme forte; les files de 0 dans cette dernière colonne caractérisent donc les formes fortes.

En ce qui concerne l'interprétation du tableau des formes fortes, nous ferons les remarques suivantes :

a) soit m le nombre total de tirages effectués, m_i le nombre d'apparitions de la $i^{\text{ième}}$ solution v^i $\left(\text{d'où } \sum_i m_i = m\right)$ et C_i l'arborescence bouclée ayant v^i pour racine; si m est suffisamment grand on peut considérer

$$\frac{m_i}{m} \neq \text{Prob}\,(x = W(v_i)) = \frac{\text{card } C_i}{\text{card } V_k};$$

si m_i est grand, on peut donc considérer que v_i est racine d'une arborescence de taille importante (card C_i grand); d'après le théorème 2, v^i est donc une solution particulièrement significative, puisque c'est un optimum local pour une grande partie de V_k.

b) Si q est le nombre de solutions obtenues, et si parmi ces solutions $v^* \in V_k$ est la solution qui minimise W, v^* est un optimum local vis-à-vis de l'ensemble des sommets des q arborescences bouclées obtenues.

c) D'après le théorème 2, les arborescences bouclées et boucles forment une partition de V_k, en conséquence plus le nombre de solutions obtenues est faible plus la taille de ces arborescences est grande et plus les solutions obtenues sont donc significatives. Dans le cas où il n'y a pas les formes le nombre d'arborescences bouclées est important et donc les solutions obtenues sont moins significatives.

d) Soit v^* la racine d'une arborescence bouclée C de $\Gamma = (V_k, h)$; soit $h^{-q}(v^*)$ le $q^{\text{ième}}$ niveau de C. Il s'avère dans la pratique que le nombre de niveaux pour une arborescence donnée est très faible : il oscille en général autour de 4 ou 5 et dépasse rarement 12, même pour des tableaux comportant 3 000 éléments à classifier. Dans le cas où il y a peu d'arborescences bouclées, le nombre de sommets d'un niveau donné est donc très grand.

EXEMPLE :

Supposons que card $(E) = 100$, $k = 15$, $n_i = 3$, le nombre de niveaux $= 5$, le nombre de solutions non-biaisées $= 6$, R est carrée. Alors la taille d'un palier est supérieure à $\dfrac{2^{100}(C_{100}^3)^{15}}{5 \times 6}$!

On a ainsi une explication de la rapidité et de l'efficacité de la méthode, qui à chaque itération permet de passer d'un niveau à l'autre en améliorant la solution.

EXEMPLES D'APPLICATIONS

7.1. L'exemple artificiel de Ruspini

Nous avons appliqué la variante *c*) sur les données de Ruspini (cf. fig. 6). Cela a permis d'abord de constater la rapidité de la méthode, par rapport à celle de Ruspini; ainsi en prenant

$$K = 4, n_1 = n_2 = n_3 = n_4 = 5,\ R(x, i, L) = D(x, C_i) = \sum_{y \in Ci} d(x, y)$$

où d est la distance Euclidienne, nous avons réalisé 50 passages de la méthode (en changeant à chaque fois le tirage de $L^{(0)}$) en 2,57 mn sur CII 10 070. Ces 50 passages ont fait ressortir l'existence de 6 arborescences bouclées. Les fréquences d'apparition de chacune des 6 solutions correspondantes sont indiquées (fig. 12). Ce graphique est en fait l'histogramme de la variable aléatoire qui a été définie en 5.1. En abcisse est représenté $U = \text{Lim } U_n$ (cf. 3.3.); la convergence est généralement atteinte au bout de 4 itérations. La solution qui apparaît le plus fréquemment est celle correspondant aux quatre meilleurs classes; la valeur de U pour cette solution est nettement meilleure que pour les autres solutions, ce qui montre qu'elle correspond bien à la meilleure partition. La meilleure solution correspond à la racine de l'arborescence bouclée de plus grande taille, ce qui est satisfaisant pour la méthode. Les solutions qui apparaissent le plus fréquemment sont indiquées (fig. 7, 8, 9, 10). On voit facilement que les solutions correspondant aux figures 9, 10 et 11 n'apportent aucune information (cf. 5.2.) à la solution donnée (fig. 7). A partir de ces solutions on obtient 4 « formes fortes » correspondant exactement aux quatre classes de la meilleure solution.

Remarquons qu'en appliquant la proposition 3 aux solutions données (fig. 8 et 9) on fait apparaître l'arborescence bouclée dont la racine est la solution correspondant à la figure 7. On donne (fig. 11) une solution obtenue en utilisant la variante du centre de gravité [cf. 3.2. *a*]; cette solution n'apparaît jamais par les variantes utilisant des noyaux car elle doit correspondre à une position instable.

On donne (tableau 1), le tableau des formes fortes, obtenu en prenant cette fois $K = 6$, $n_5 = n_6 = 5$, sans changer les autres paramètres et en réalisant 5 passages de la méthode ($n = 5$). Ce tableau fait ressortir l'existence de 6 formes fortes et 3 formes faibles. Notons, B_1, B_2, B_3 les formes faibles et A_i

les formes fortes (cf. fig. 13). On peut mesurer la « faiblesse » de B_i en utilisant la fonction F (cf. 5.2.2.). Comme

$$\sum_{j=1}^{3} h_{A_j}(x) = 1 + \frac{4}{5} + \frac{4}{5} \forall\, x \in B_1 \;,\; \sum_{j=4}^{5} h_{A_j}(x) = 1 + \frac{1}{5} \forall\, x \in B_2$$

et
$$h_{A_6}(x) = 1 \;\forall\, x \in B_3$$

on a :

$$F(B_1) = \frac{13}{15},\; F(B_2) = \frac{3}{5} \text{ et } F(B_3) = 1.$$

On voit que B_3 est une forme forte, que B_1 est presque une forme forte et que B_2 est une forme relativement faible. Ces formes fortes et faibles telles qu'elles apparaissent (fig. 13) expriment bien ces valeurs. Signalons enfin que dans 4 des 5 solutions il apparaît des classes vides ce qui signifie que le nombre de classes existant réellement doit être plus petit que 6.

7.2. Classement de sondages d'un gisement minier

L'étude dont il est question a été réalisée par J. Picard pour sa thèse de 3e cycle (cf. [17]). Elle porte sur 149 sondages géologiques. Chaque sondage est caractérisé par 24 teneurs métal mesurées de mètre en mètre sur une profondeur totale de 24 mètres. Les diverses méthodes d'analyse de données utilisées (analyse factorielle des correspondances et en composantes principales) n'ont pas permis de distinguer les différents types de courbes teneur métal/profondeur, contenues dans la population.

J. Picard a alors utilisé la méthode des nuées dynamiques d'une part pour une classification à l'aide d'étalons initiaux choisis, d'autre part pour une recherche de profils types à l'aide des formes fortes. L'enrichissement apporté aux méthodes classiques peut être résumé comme suit :

1) Le procédé de classification permet un partitionnement du plan 1-2 de l'analyse factorielle non décelable à priori, ce découpage en éléments disjoints est confirmé encore plus nettement dans l'espace tridimensionnel.

2) Le tableau des formes fortes fait apparaître 5 formes fortes dont les sondages moyens sont très significatifs d'un type de terrain.

3) Les 5 formes fortes réunies ne contiennent que 59 sondages. J. Picard a vérifié la représentativité de ces 5 formes en montrant que le nuage obtenu par les deux analyses factorielles (1) suivantes ne présentait pas de modification sensible :

— une analyse factorielle des correspondances des formes fortes avec en éléments supplémentaires les sondages restants,

— une analyse factorielle des correspondances de toute la population.

(1) Ces analyses factorielles sont représentées dans la thèse de J. Picard.

Une autre expérience a permis de confirmer ce résultat : la position des paramètres dans une analyse factorielle sur la population totale est la même que dans une analyse factorielle faite uniquement sur les 59 sondages des formes fortes.

4) Il a en outre été procédé à une classification hiérarchique qui ne contredisait pas les résultats obtenus mais ne faisait pas apparaître nettement les formes fortes.

On peut remarquer que les deux méthodes peuvent être utilisées conjointement, la méthode hiérarchique permettant une évaluation du nombre de classes a priori pour la méthode des nuées dynamiques, cette dernière déterminant des classes très typées et les profils principaux de la population.

7.3. Recherche de profils biologiques (1)

Il s'agit de découvrir des groupements types dans une population de 990 sujets. Nous avons tenu compte de 16 paramètres chez chaque malade : des proportions des 5 fractions électrophorétiques : albumine, alpha-1, alpha-2, bêta, gamma, des trois paramètres de la fiche réticulo-endothéliale et du taux de 7 globulines individuelles, déterminées par la méthode immunochimique de diffusion radiale.

Ces 900 sujets ont été classés a priori en 33 groupes distincts représentant soit des entités nosologiques, soit des syndromes.

Sont représentés en particulier, le cancer des tissus solides sans ou avec atteinte du foie, les leuco-réticuloses, des maladies infectueuses très variées, des collagénoses, la cirrhose du foie et l'hépatite virale, des dermatoses variées, des maladies atopiques, le diabète et autres troubles endocriniens, la macroglobulinémie des africains, des ulcères gastro-duodénaux. Un travail approfondi sur les mêmes données a été réalisé par le P^{r} Lenoir et M. Kerbaol à l'aide de l'analyse factorielle des correspondances (cf. [21]; cette analyse a été faite sur un nombre restreint de paramètres et a fait apparaître des nuages d'un grand intérêt pour les praticiens; cependant, une délimitation objective de ces nuages est difficile et de plus sur les 16 paramètres avec 990 sujets le nuage obtenu par l'analyse factorielle est d'interprétation difficile; d'un autre côté, une classification donnant une hiérarchie n'est pas praticable vu la taille des données.

Nous avons utilisé la variante c) de la méthode des nuées dynamiques avec $k = 10$, $n_i = 10$ et la distance du χ^2 (cf. [10]). Le tableau des formes fortes a été calculé avec $n = 15$; les formes fortes obtenues sont particulièrement

(1) Ce travail a été réalisé en collaboration avec le M. le Professeur Sandor de l'Institut Pasteur, MM. Lechevalier et Barré de l'IRIA; il a fait l'objet d'une communication à l'Académie des Sciences [22], et d'un rapport de stage IRIA, pour plus de détails le lecteur pourra se reporter à ces textes.

significatives puisque chaque sujet a une chance sur 10^{15} d'appartenir à une forme forte donnée.

Le Pr Sandor a trouvé très commode la représentation en nombres entiers (donnée par le tableau des formes fortes) pour exprimer la position réelle des sujets dans $\mathbf{R}^{16}$. En effet, grâce à ce tableau on a un moyen de saisir les multiples aspects de la position des points dans $\mathbf{R}^{16}$, de manière bien plus proche de la réalité que toute classification rigide n'aurait pu le faire.

Une étude détaillée du tableau des formes fortes et une analyse factorielle du triple (cf. [4]) sur les 51 formes fortes qui sont apparues ont permis de dégager nettement l'existence de huit formes, ces formes permettent de tracer 8 profils types.

Nous ne donnerons pas ici l'interprétation détaillée des profils types obtenus, le lecteur intéressé pourra se reporter au compte rendu à l'Académie de Médecine (séance du 29-2-1972) à paraître prochainement. Nous nous bornerons à donner la conclusion de compte rendu.

Après avoir signalé une anomalie en ce qui concerne le classement de l'ataxie télangiéctasique (1), le Pr Sandor conclut ainsi : « Il reste non moins vrai que les résultats que nous apportons constituent une excellente base d'un diagnostic objectif. L'appartenance à un type de profil donnera, en effet, le plus souvent tous les renseignements que le praticien peut tirer sur le plan des diagnostics et des pronostics d'un protéinogramme et il n'aura pour cela besoin d'aucune connaissance concernant la nature et l'origine des diverses protéines sériques ».

CONCLUSION

Un large champ de recherche reste ouvert; sur le plan pratique, il faudrait développer à l'aide des méthodes d'apprentissage par exemple, le choix de f et g, développer les techniques permettant le choix de k (le nombre de classes demandées a priori), réaliser une comparaison exhaustive des différentes variantes de la méthode des nuées dynamiques, approfondir et développer les techniques du passage d'une arborescence à l'autre, faire une étude statistique de la structure de l'espace V_k en liaison avec E, notamment en ce qui concerne le nombre relatif d'éléments impasses, d'éléments non biaisés, la taille des arborescences, des niveaux, etc... Mettre au point des techniques permettant une vision plus nette du tableau des formes fortes (du type « minimum spaning tree » par exemple). Utiliser les formes faibles afin de détecter entre les formes fortes les zones à faible densité (l'obtention des « trous » débouchant sur de nombreuses applications pratiques).

(1) Nous pensons que cette anomalie vient du fait que la distance utilisée donne plus d'importance aux augmentations au-dessus de 1 qu'aux diminutions entre 0 et 1.

Sur le plan théorique il faudrait caractériser les familles de fonctions carrées, développer des théorèmes de convergence pour les différentes variantes et dans le cas où le nombre d'objets à classifier tend vers l'infini, ce dernier point est d'un grand intérêt pratique car il devrait permettre de développer et justifier des techniques purement séquentielles.

Remerciements. — Je tiens à témoigner ma reconnaissance à M. le Professeur J. C. Simon (1) pour ses conseils et ses encouragements ainsi qu'à M. Chavent (2) pour ses judicieuses remarques lors de la rédaction définitive de ce texte. Également MM. M. Roux (3) pour ses conseils, Y. Lechevallier (3) et J. Barré (3) pour leurs remarques et leur aide à la programmation.

BIBLIOGRAPHIE

[1] BALL G. H., *Classification Analysis*, Technical Note, Stanford Research Institute. Menlo Park, California 94025 USA 1970.

[2] BARBU M., *Partitions d'un ensemble fini: leur treillis*, M.S.H. n° 22, 1968.

[3] BENZECRI J. P., *Algorithmes rapides d'agrégation*, Sup. Class. n° 9, Laboratoire de Statistique Mathématique, Université de Paris-6, 1971.

[4] BENZECRI J. P., *Représentation Euclidienne d'un ensemble muni de masses et de distances*, Université de Paris-6, 1970.

[5] BERGE C., *Théorie des graphes et ses applications*, Dunod Éditeur, Paris, 1967. BOLSHEV L. N., *Cluster Analysis*, I.S.I.R.S.S.' 69, 1969.

[6] BONNER R. E., *On some clustering technics*, IBM Journal of Research and Development, 1964.

[7] CORMACK R. M., *A review of Classification*, The journal of the Royal Statistical Society, Serie A, 1971, vol. 134, Part 3.

[8] DIDAY E., BERGONT M., BARRÉ J., *Différentes notes sur la programmation de la Méthode des nuées dynamiques*, Note IRIA, Rocquencourt 78, 1970-71-72.

[9] DIDAY E., *La méthode des nuées dynamiques et la reconnaissance des formes*, Cahiers de l'IRIA, Rocquencourt 78, 1970.

[10] DIDAY E., *Une nouvelle méthode en classification automatique et reconnaissance des formes*, Revue de Statistique Appliquée, 1971, vol. XIX, n° 2.

[11] FISHER L., VAN NESS J. W., *Admissible Clustering Procedures*, Biometrika, 1971, 58, 1, p. 91.

[12] FREEMAN N., *Experiments in discrimination and classification*, Pattern Recognition J., 1969, vol. 1, n° 3.

[13] HALL D. J. BALL G. H., *Isodata a Novel Method of Data Analysis and Pattern Classification*, Technical Report, 5 R I Project 5533. Stanford Research Institute, Menlo Park, California U.S.A., 1965.

[14] HILL D. R., *Mechanized Information Storage, retrevial and dissemination*, Proceedings of the F.I.D./I.F.I.P. Joint Conference Rome, 1967.

[15] JOHNSON S. C., *Hierarchical clustering schemes*, Psychometrica, 1967, 32, 241-45.

---

(1) Professeur à la Faculté des Sciences de Paris-6.
(2) Chef de projet à l'IRIA dans le département de M. le Professeur Lions.
(3) Laboratoire de Statistique mathématique de la Faculté des Sciences de Paris (dirigé par M. le Professeur J. P. Benzecri).

[16] LERMAN H., *Les bases de la classification automatique*, Gauthiers-Villars, 1970.

[17] PICARD J., *Utilisation des méthodes d'analyse de données dans l'étude de courbes expérimentales*, Thèse de 3e cycle. Laboratoire de Statistique Mathématique, Université Paris 6, 1972.

[18] ROMEDER J. M., *Méthodes de discrimination*, Thèse de 3e cycle. Statistique Mathématique. Faculté des Sciences de Paris 6, 1969.

[19] ROUX M., *Un algorithme pour construire une hiérarchie particulière*, Thèse de 3e cycle. Laboratoire de Statistique Mathématique, Université de Paris 6, 1968.

[20] RUSPINI H. R., *Numerical Methods for fuzzy clustering*, Information Science 1970, 2, p. 319-350.

[21] SANDOR G., LENOIR P., KERBAOL M., *Une étude en ordinateur des corrélations entre les modifications des protéines sériques en pathologie humaine*, C. R. Acad. Sc. Paris, 1971, 272, p. 331-334.

[22] SANDOR G., DIDAY E., LECHEVALLIER Y., BARRÉ J., *Une étude informatique des corrélations entre les modifications des protéines sériques en pathologie humaine*, C. R. Acad. Sc. Paris, 1972, t. 274, d.p. 464-467.

[23] SEBESTIEN G. S., *Automatic off-line Multivariate Data Analysis*, Proc. Fall Joint Computer Conference, 1966 pp. 685-694.

[24] SOKHAL R. R., SNEATH P. H. R., *Numerical Taxonomy*, W. H. Freeman and Co., San Francisco and London, 1963.

[25] WATANABÉ M. S., *A unified view of clustering algoritms*, IFIP Congress 71, Ijubiana, Booklet TA-2, 1971.

[26] ZADEH L. A., *Fuzzy sets*, Inf. Control, 1965, 8, pp. 338-353.

[27] ZAHN C. I., *Graph theoretical methods for detecting and describing Gestalt Clusters*, I.E.E.E. Trans. on Computers, vol. C-20, n° 1, January, 1971.

[28] McQUEEN J., *Some Methods for Classification and Analysis of Multivariate Observations*, 5th Berkeley Symposium on Mathematics, statistics and probability, 1967, vol. 1, n° 1, pp. 281-297.

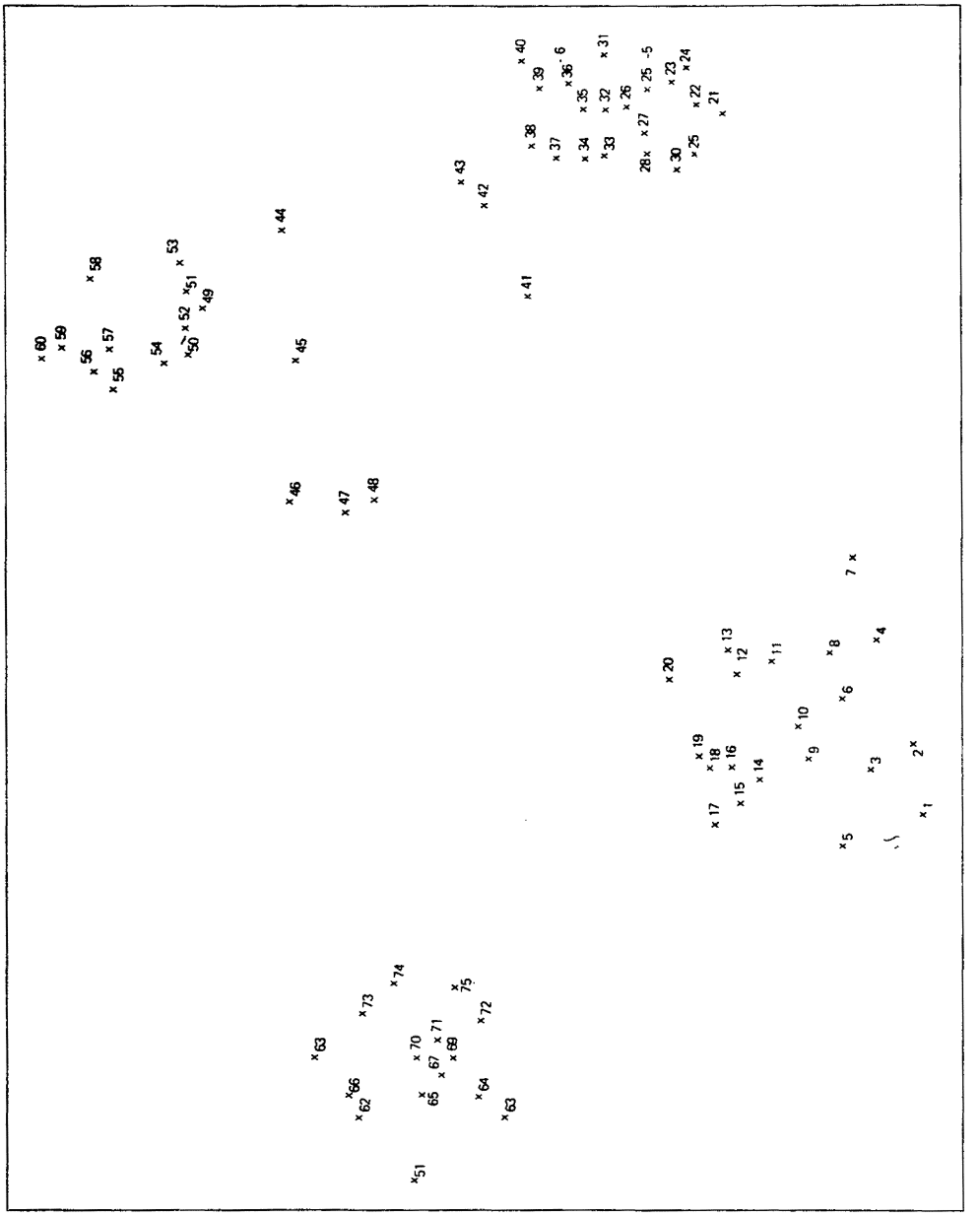


Figure 6

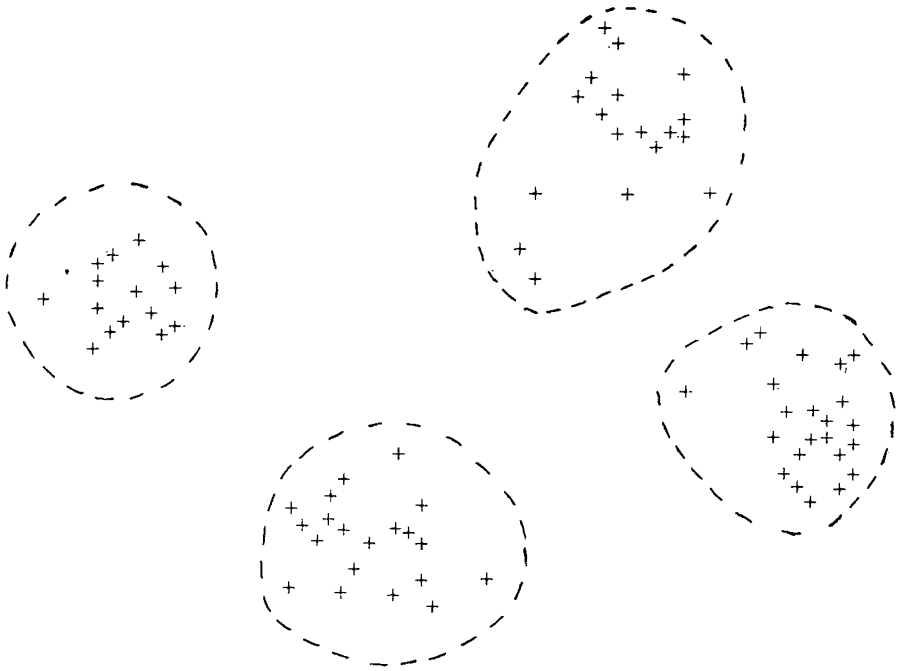


Figure 7

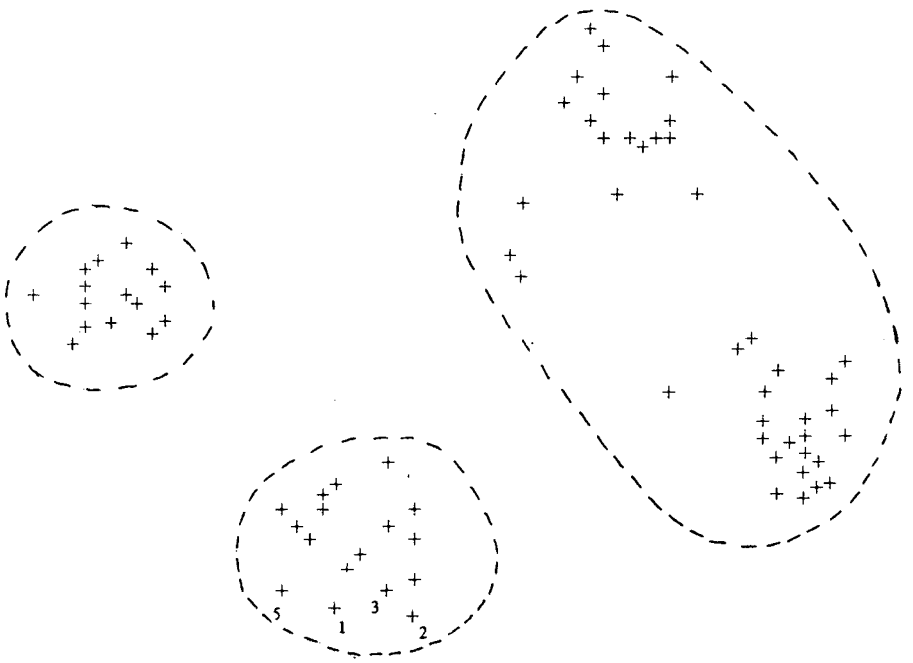


Figure 8

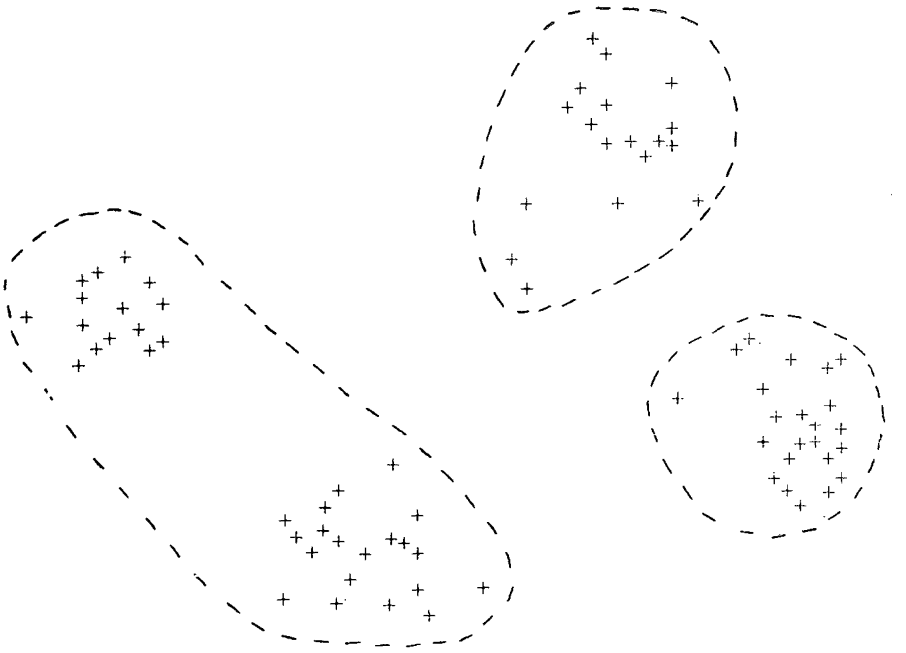


Figure 9

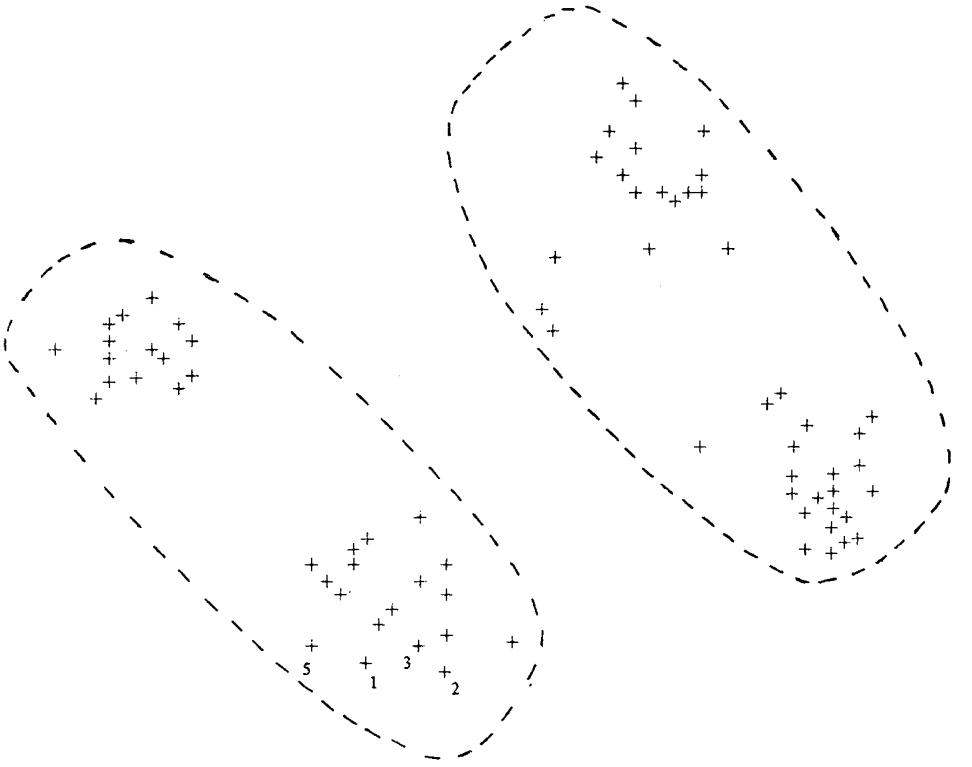


Figure 10

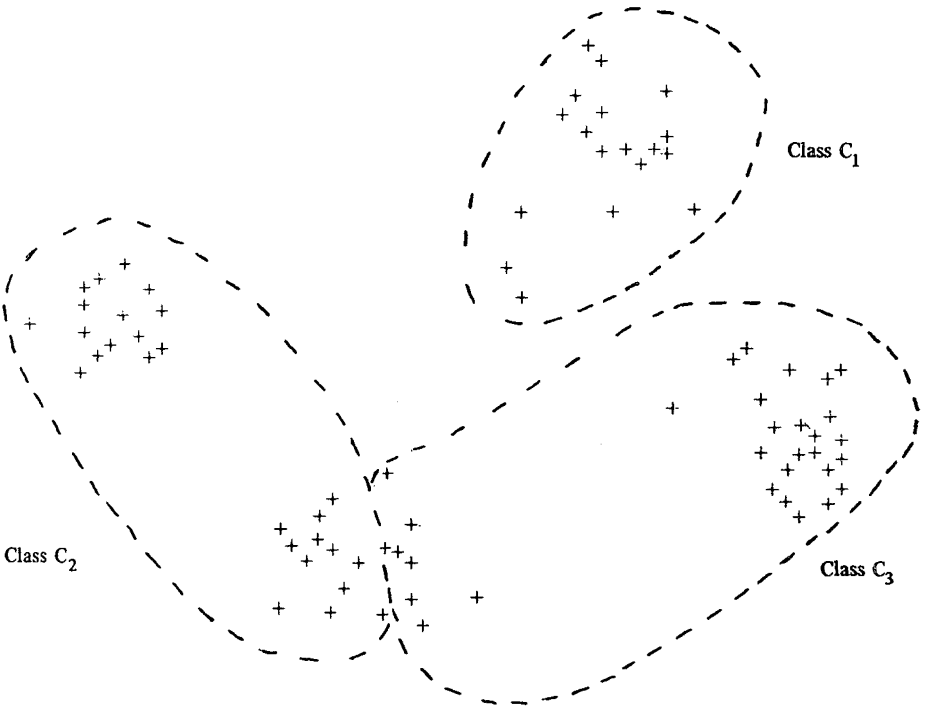


Figure 11

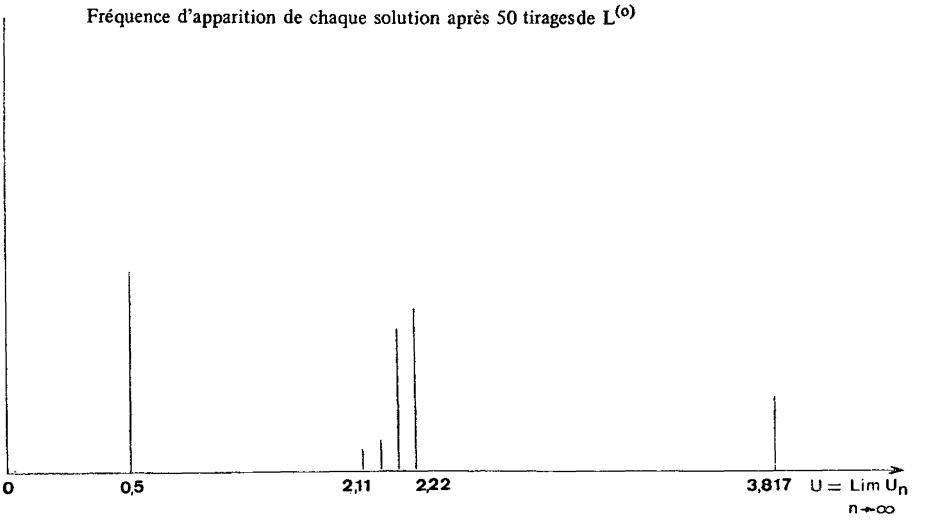


Figure 12

Fréquence d'apparition de chaque solution après 50 tirages de $L^{(o)}$.

La valeur $U = 0,5$ correspond à la solution donnée (fig. 7).
$U = 2,11$ correspond à la solution donnée (fig. 8).
$U = 2,22$ correspond à la solution donnée (fig. 9).
$U = 3,817$ correspond à la solution donnée (fig. 10).

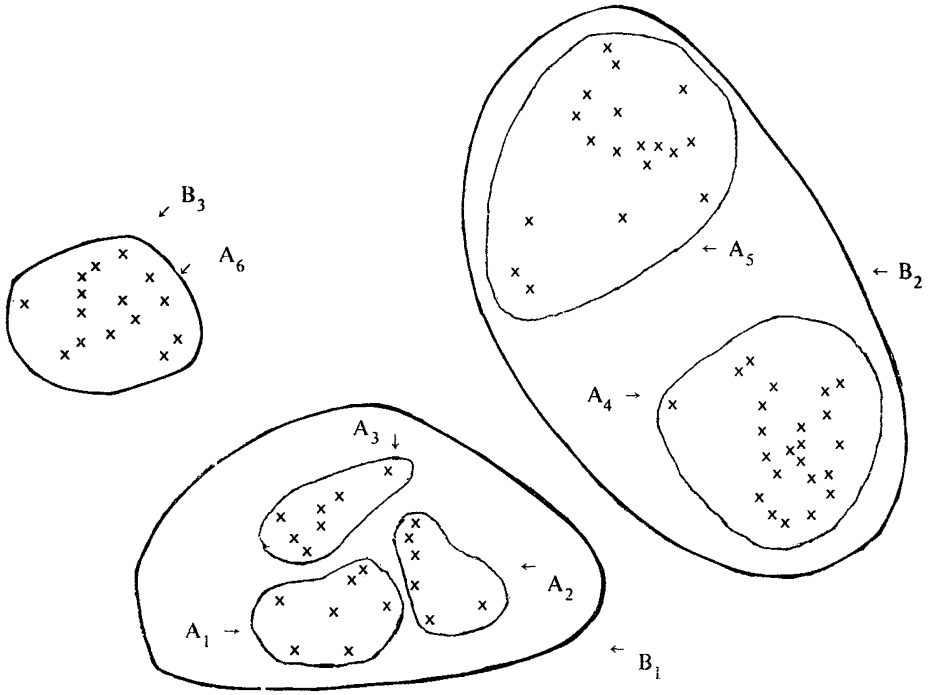


Figure 13

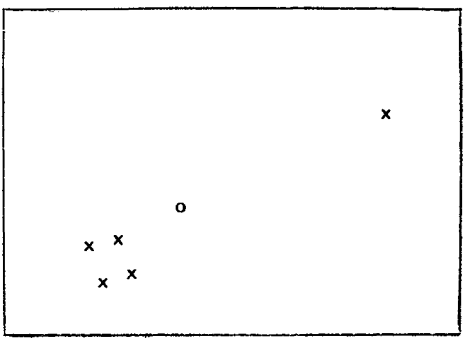


Figure 14

Les signes « × » représentent les éléments à classifier alors que le signe « o » représente le centre de gravité des 5 éléments.

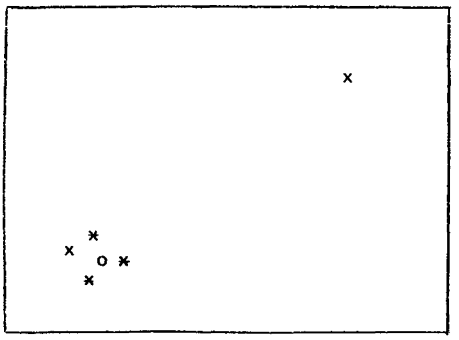


Figure 15

Les 3 éléments les plus proches de la population sont représentés par le signe « * » ; le centre de gravité de ces 3 éléments atténue l'effet du point marginal.

ANNEXES

ANNEXE 1

Soit B un ensemble fini et une fonction $h : B \to B$. Le graphe défini par B et l'ensemble des arcs $(h(x), x)$ sera noté $\Gamma = (B, h)$. On sait que l'ensemble des composantes connexes de Γ constitue une partition de B; chacune de ces composantes à une forme particulière :

Proposition 1 :

Chaque composante connexe de Γ contient un circuit au maximum.

Démonstration :

Si dans une composante connexe il existe un circuit, cela implique l'existence d'une suite finie de sommets $C = \{x_0, ..., x_n\}$ telle que $h(x_i) = x_{i+1}$ et $h(x_n) = x_0$. L'existence de deux circuits dans une composante connexe implique la possibilité de sortir d'un circuit, c'est-à-dire l'existence de i et $y \notin C$ tels que $h(x_i) = y$. Cela n'est pas possible puisque $h(x_i) = x_{i+1}$ et que h est une fonction.

c.q.f.d.

On donne (fig. 4) un exemple de composante connexe de Γ. Nous dirons que x est un point fixe si $x = h(x)$. Une arborescence ayant pour racine un point fixe sera appelée arborescence bouclée (voir fig. 5). Soit W une application $B \to \mathbf{R}^+$.

Proposition 2 :

Si W est injective sur toute suite v_n et vérifie la propriété $W(h(x)) \leqslant W(x)$ alors :

1) Chaque composante connexe de Γ contient une boucle et une seule et ne contient pas un autre circuit.

2) Chaque composante connexe de Γ est une arborescence bouclée ou une boucle.

3) Si $y \in B$ n'est pas un point fixe, il existe un point fixe x tel que $W(x) < W(y)$.

Démonstration :

1) Soit un circuit $C = \{x_0, ..., x_n\}$ où $h(x_i) = x_{i-1}$, la condition $W[h(x)] \leqslant x$ implique $W(x_0) \leqslant W(x_1) \leqslant ... \leqslant W(x_n) \leqslant W(x_0)$ d'où $W(x_0) = W(x_i)$ $\forall_i = 1, 2, ..., n$.

Comme W est injective on a : $x_0 = x_i$. Donc tout circuit de Γ est une boucle.

Soit un chemin $C = \{x_0, ..., x_n\}$, la suite $u_n = W(x_n)$ est décroissante et minorée par 0; elle converge donc, de plus elle atteint sa limite (cf. [15]). Il existe donc N : $\forall n \leqslant N u_n = u_{n+1}$, d'où $x_n = x_{n+1}$ d'où $h(x_n) = x_n$. Donc toute partie connexe de Γ contient une boucle et une seule, elle ne contient pas un autre circuit d'après la proposition 1.

2) D'après ce qui vient d'être prouvé, toute composante connexe a un sommet et un seul qui est un point fixe. Soit x_1 ce sommet. Soit $X = \{x \in B / h(x) = x_1, x \neq x_1\}$. Si $X = \emptyset$, la composante connexe contenant x_1 est réduite à une boucle. Si $X \neq \emptyset$, la composante connexe contenant x_1 est une arborescence bouclée; en effet, si on supprime la boucle $(h(x_1), x_1)$, les trois propriétés définissant une arborescence de racine x_1 sont vérifiée (1) :

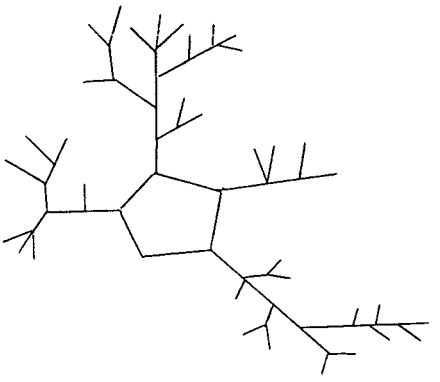
— Tout sommet $\neq x_1$ est l'extrêmité terminale d'un seul arc; cela vient du fait que h est une fonction.

— x_1 n'est l'extrémité terminale d'aucun arc; puisqu'on a supprimé la boucle $(h(x_1), x_1)$.

— La composante connexe contenant x_1 n'a pas de circuit; cela d'après la première assertion de cette proposition.

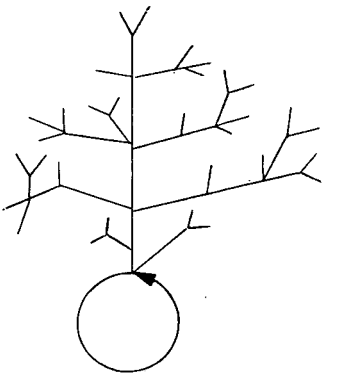
3) Les composantes connexes de Γ constituent une partition (2) de B; si $y \in B$, y est un sommet d'une composante connexe de Γ; si y n'est pas un point fixe, il appartient à une arborescence bouclée. Soit x_1 la racine de cette arborescence; il existe un chemin de x_1 à y, soit $C = \{x_1, x_2, ..., x_n\}$ ce chemin, où $x_n = y$. On a : $W(x_1) \leqslant W(x_2) \leqslant ... \leqslant W(x_n)$ d'où $W(y) \geqslant W(x_1)$. Comme W est injective et y n'est pas un point fixe on en déduit $W(y) \geqslant W(x_1)$.

c.q.f.d.



Composante connexe de Γ

Figure 4



Arborescence bouclée

Figure 5

(1) Voir [5], page 154.
(2) Voir [5], page 9.

ANNEXE 2

Il s'agit de montrer que les deux propriétés suivantes sont équivalentes pour caractériser une forme faible.

1) Q^* est la plus fine des partitions qui sont moins fines que $P^{1*}, ..., P^{n*}$.

2) Q^* est la partition définie par l'ensemble des parties connexes du graphe $\Gamma_1 = (E, F_1)$.

Démonstration :

Nous allons d'abord montrer que 1) $\Rightarrow$ 2). Soit $x \in Q_j^*$ alors

$$\forall z \in \complement_E Q_j^* \text{ et si } W(z) = (\gamma_1, ... ; \gamma_n)$$

on a $\alpha_j - \gamma_j = o$ car sinon Q^* ne serait pas moins fine que P^{i*} $\forall$ i. Cela revient à dire que $\forall z \in \complement_E Q_j^*$ et $\forall x \in Q_j^*$ on a $z \notin F_1(x)$. Pour montrer que Q_j^* est bien une partie connexe de Γ_1, il reste à vérifier que :

$$\forall x \in Q_j^*, \exists y \in Q_j^* : y \in F_1(x).$$

Prenons x quelconque dans Q_j^* et supposons qu'il n'existe pas y appartenant à Q_j^* tel que $y \in F_1(x)$; cela signifierait que $\delta(x, y) = 0$ $\forall$ $y \in Q_j^*$ autrement dit $\alpha_i \neq \gamma_i \forall_i$; ainsi en remplaçant Q_j^* par les classes $Q_j - \{x\}$ et $\{x\}$ on définit une partition Q qui tout en étant moins fine que les P^{i*} serait plus fine que Q^* ce qui est contraire à l'hypothèse.

Montrons maintenant que 2) $\Rightarrow$ 1). Par définition de Γ_1 une partie connexe Q_j est telle que $\forall x \in Q_j$ et $\forall z \in \complement_E Q_j$ on a : $z \notin F_1(x)$, c'est-à-dire $\alpha_i - \gamma_i = o \forall i$; ainsi l'ensemble des parties connexes Q_j de Γ_1 constitue bien une partition moins fine que les partitions P^{i*} $\forall$ i. Il reste à montrer que la partition Q des parties connexes de Γ_1 est bien la plus fine parmi celles qui sont moins fines que les P^{i*} ; soit x et y deux éléments appartenant à la même partie connexe de Γ_1 ; il existe une chaîne $z_1, ..., z_n$ tel que $x = z_1$, $y = z_n$ et $z_{q+1} \in F_1(z_q) \forall q = 1, 2, ..., n - 1$; deux sommets consécutifs quelconques de cette chaîne appartiennent nécessairement à une même classe de l'une des partitions P^{i*}, par définition même de F_1. Pour toute partition plus fine Q' que Q qui découpe par exemple Q_j en deux parties A et B il existe $x \in A$ et $y \in B$ et donc $q : z_q \in A$ et $z_{q+1} \in B$; Q' découpe donc l'une des classes de la partition P^{i*} et n'est donc pas moins fine que les partitions P^{i*} $\forall$ i. La plus fine de ces partitions est donc bien Q.

c.q.f.d.

ANNEXE 3

Théorème (des « connexités descendantes »)

Soit Δ l'application [1] $ExE \rightarrow \mathbf{N}$ telle que $\Delta(x, y) = n - \delta(x, y)$ et soit E' l'espace quotient [2] E/H. Si F_p est la multi-application $E' \rightarrow \mathbf{P}(E')$ telle que $F_p(x) = \{ y \in E' / \delta(x, y) \geqslant p \}$ et Γ_p est le graphe (E', F_p), alors :

1) L'ensemble des parties connexes de Γ_p pour $p = 0, 1, 2, ..., n$ constitue une hiérarchie sur E'.

2) Cette hiérarchie induit l'ultramétrique sous-dominante de Δ.

Démonstration

1) Soit G_i^p la $i^{\text{ième}}$ classe de la partition définie par les parties connexes du graphe Γ_p.

Soit $G = \{ G_i^p / i = 1, ..., q_p ; p = 1, ..., n \}$ où q_p est le nombre de parties connexes de Γ_p; nous allons montrer que G est une hiérarchie sur E' :

— $E \in G$, car Γ_0 est réduit à une seule partie connexe qui est identique à E.

— $\forall x \in E'$ on a $x \in G$; en effet, chaque partie connexe de Γ_n est réduite à un seul élément et l'ensemble de ces parties constitue une partition de E'.

— Quels que soient a et b, éléments de G, si $a \cap b \neq \varnothing$ alors on a soit $a \subset b$, soit $b \subset a$. En effet, posons $a = G_i^p$, $b = G_i^m$, deux cas peuvent se produire : $p = m$, alors $a \cap b = \varnothing$ puisque a est une partie connexe et b une autre partie connexe du même graphe Γ_p.

$p > m$, soit $x \in G_i^p$ alors pour tout élément $y \in G_i^p$ il existe une chaîne $z = z_1, ..., z_q$ avec $z_1 = x$ et $z_q = y$ telle que $\text{Min } \delta(z_i, z_{i+1}) \geqslant p > m$; donc tous les éléments connexes à x dans Γ_p sont dans une même partie connexe de Γ_m, autrement dit G_i^p est contenu dans une des parties connexes de Γ_m, on a donc $G_i^p \subset G_j^m$ ou bien $G_i^p \cap G_j^m = \varnothing$.

Ainsi G est une hiérarchie.

2) On peut indicer G par l'application $X : G \rightarrow [0, 1]$ telle que $X(a) = n$ si p est le plus grand entier tel que $a \equiv G_i^p$. Cette application définit bien une hiérarchie indicée puisque $x(a) = 1$ si a est réduit à un seul élément car alors il existe $i : a \equiv G_i^n$; d'une part $a \subset b \Rightarrow X(a) > X(b)$ car si $a = G_i^p$ et $b = G_j^m$, on déduit de 1) que $a \subset b \Rightarrow p > m$ d'où $X(a) > X(b)$.

---

(1) Δ est en fait la métrique de la différence symétrique (voir définition de δ en 5.2.1.).
(2) H est l'application qui a été définie en 5.2.1.

La hiérarchie G ainsi indicée, permet de définir sur E' un indice de similarité de la manière suivante :

$$d(x, y) = n(1 - \operatorname*{Max}_{a} \{ X(a)/x, y \in a \}) ;$$

autrement dit $d(x, y) = n - q$ où q est le plus grand entier tel que x et y appartiennent à la même partie connexe de Γ_q. Cela implique l'existence d'une chaîne $z = (z_1, ..., z_p)$ telle que $z_1 = x$, $z_p = y$ et

$$\inf_{z \in \mathcal{C}_{xy}} \operatorname*{Max}_{i} (n - \delta(z_i, z_{i+1})) = n - q$$

où $\mathcal{C}_{xy}$ est l'ensemble des chaînes de x à y. En effet, dire que x et y appartiennent à une même partie connexe de Γ_q signifie qu'il existe une chaîne z telle que $\operatorname{Min} \delta(z_i, z_{i+1}) \geqslant q$ d'où :

$\operatorname*{Max}_{z \in \mathcal{C}_{xy}} \operatorname*{Min}_{i} \delta(z_i, z_{i+1}) \geqslant q$; si $\operatorname*{Max}_{z \in \mathcal{C}_{xy}} \operatorname*{Min}_{i} \delta(z_i, z_{i+1}) \neq q$ on a pour tout $z \in \mathcal{C}_{xy}$ $\operatorname*{Min}_{i} \delta(z_i, z_{i+1}) \geqslant q + 1 > q$ ce qui est contraire au choix de q qui est le plus grand des entiers tel que x et y appartiennent à une même partie connexe de Γ_q. On a donc $\operatorname*{Max}_{z \in \mathcal{C}_{xy}} \operatorname*{Min}_{i} \delta(z_i, z_{i+1}) = q$ d'où

$$\operatorname*{Min}_{z \in \mathcal{C}_{xy}} \operatorname*{Max}_{i} (n - \delta(z_i, z_{i+1}) = n - q,$$

ce qui implique : $d(x, y) = \operatorname*{Min}_{z \in \mathcal{C}_{xy}} \operatorname*{Max}_{i} \Delta(z_i, z_{i+1})$. Cette condition suffit à prouver que d est la sous-dominante de Δ (cf. M. Roux [19]).

c.q.f.d.

REMARQUE :

Ce théorème permet d'obtenir un bon rangement du tableau des formes fortes. Plus généralement, pour tous les problèmes où le tableau des données ne contient que des nombres entiers et pour lesquels la distance de la différence symétrique est significative, ce théorème donne une méthode commode pour la construction de la hiérarchie indicée induite par la sous-dominante. En effet, on économise du temps et de la place mémoire puisqu'on peut construire cette hiérarchie sans avoir besoin de calculer et de mettre en mémoire le tableau des distances deux à deux des éléments pour la différence symétrique et pour la distance associée à la sous-dominante.